DOCUMENT RESUME

ED 166 197 TM 007 533

AUTHOR Barrows, Thomas S.; And Others
TITLE Psychoeducational Diagnostic Services for Learning Disabled Youths. Research Procedures.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY National Inst. for Juvenile Justice and Delinquency Prevention (Dept. of Justice/LEAA), Washington, D.C.
PUB DATE 77
GRANT 76-99-0022; 76NI-99-0133
NOTE 39p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Clinical Diagnosis; Delinquents; *Diagnostic Tests; Emotionally Disturbed; Intelligence Tests; *Learning Disabilities; Males; Mentally Handicapped; *Psychological Testing; Research Methodology; Secondary Education

ABSTRACT

A detailed explicit definition of learning disability (LD) is presented in terms of the diagnostic procedures and decision points used in identifying that category among a group of juvenile delinquent and control group adolescent males. The described procedures include a review of existing records, interviews, and diagnostic assessments. Rationales for the choice of pertinent variables and measurement instruments are given. The assessment process and administrative arrangements are explained. The factor analytic studies of Witkin and his colleagues were deemed especially relevant and the Wechsler Intelligence Scale for Children (Revised) was used. Decision points eliminated boys who were mentally retarded (intelligence quotient below 75), and boys who were, or had been, diagnosed as emotionally disturbed (those diagnosed as untestable or presenting bizarre symptoms). Those chosen for further diagnosis included boys whose achievement score differed by two or more years from their ability score, or boys whose academic achievement was two or more years retarded. Interviews and diagnostic testing were carefully standardized and required about three hours for each boy. Training, quality control, and other administrative procedures are described. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED166197

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Psychoeducational Diagnostic Services for Learning Disabled Youths

Research Procedures

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. Weidemiller
Educational Testing Service

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Thomas S. Barrows
Paul B. Campbell
Billie A. Slaughter
Mary Louise Trainor

TM007 533

Copyright © 1977 by Educational Testing Service. All rights reserved.

Acknowledgements

A document of this complexity of necessity represents the thinking of many people. In addition to the authors, other substantial contributions came from able consultants who reviewed our plans and provided helpful suggestions. Dick Harsh and Helen Hursch both provided insights from their experience in dealing with delinquent boys. Estelle Jasnoff presented helpful suggestions in scoring and on evaluating the effect of medication. Ronald Finkenbinder and Joanna Williams both suggested clarifications which have improved the presentation. Margaret Jo Shepherd and Suzanne Zunzer added many useful suggestions for interpreting the results of the assessments. The analysis and interpretation of Paul Broder, Dorothy Crawford, Charles Wheeler, Georgia Yanicke and Joel Zimmerman were of major value. Each of these people also made suggestions and raised questions with which we were unable to deal under the restraints of time and budgets, but which were nevertheless insightful and clarifying.

Finally we wish to commend Sue Fesmire, Donna Fleig and Helen Tarr for their secretarial skills, patience and ability to read illegible handwriting. They have done a remarkable job in keeping the project on schedule.

PBC

Prepared under Grants 76NI-99-0133 and 76-99-0022 from the National Institute of Juvenile Justice and Delinquency Prevention, Law Enforcement Assistance Administration, U. S. Department of Justice.

Points of view or opinions in this document are those of the authors and do not necessarily represent the official position or policies of the U. S. Department of Justice or Creighton University.

Foreword

The following document outlines the diagnostic procedures that are being employed in a research and demonstration program that is investigating the relationship between specific learning disabilities and juvenile delinquency. It is hoped that this documentation of the procedures that were followed will be useful to researchers, clinicians and others who are concerned with making determinations regarding the presence of learning disabilities (LD) in adolescents.

The two-year program was funded in October, 1976, by the National Institute for Juvenile Justice and Delinquency Prevention of the Law Enforcement Assistance Administration through grants to the Association for Children with Learning Disabilities (ACLD) and Creighton University's Institute for Business, Law and Social Research. Creighton University contracted with Educational Testing Service to perform the diagnostic evaluations of the participants in the study. The program has three major components: (1) the determination of the prevalence of LD in groups of adjudicated delinquents and officially nondelinquent 12-to-15 year old boys; (2) a remediation program for selected groups of adjudicated delinquents who are judged LD; (3) and an evaluation of the effectiveness of the remediation program.

The program is being conducted primarily in Phoenix, Arizona; Indianapolis, Indiana; and Baltimore, Maryland. The incidence study and the evaluation are being conducted by the Creighton Institute with Educational Testing Service, while the remediation program is the responsibility of the ACLD.

The incidence study will provide data systematically concerning the prevalence of LD in adolescent males, and will be used to select participants for the remediation program. An operational definition of LD has been adopted for the study, and procedures and criteria for making diagnostic judgments have been specified. The explication of the definition and the decision process is the focus of this paper. Preliminary results of the incidence study will be released early in 1978.

We believe that this effort is unique in the degree to which the criteria for making diagnostic decisions have been specified explicitly. We realize that no single document can be the "final word" in any field as dynamic as the study of learning disabilities; however, we hope that this documentation will be useful in future efforts.

Paul K. Broder
Creighton University
July, 1977

Table of Contents

Page

I. Overview 1

II. Theoretical Approach to the Assessment of Learning Disabilities 4

- Variable selection
- Instrument selection
- Diagnostic decisions and incidence estimates

III. Process 12

- Records review
- Interviews
- Diagnostic assessment

IV. Administrative Arrangements. 23

- Major Tasks and Staffing
- Training
- Quality control
- Data forms

V. Application 28

Figures I through IV 29-33

References 34

I. OVERVIEW

The task addressed by these procedures can best be considered in several parts. In sequence, they include reviewing records and applying decision rules for the selection of students for further diagnostic assessments, conducting diagnostic assessment, making recommendations for remediation, initial analysis and post-remediation data collection. Paralleling these major procedures are several activities, including the briefing of assessors, quality control, and information reporting to several audiences.

There were certain constraints which were operating during the implementation of the study which must be kept in mind when one considers the selected procedures. Two of them have had significant, and to some extent, controlling influence on the procedures design. First is the nebulousness of and lack of agreement on current definitions of learning disability. The proposed rules (Federal Register, 1976) governing the allocation of federal assistance to states, for example, include the following comments:

1. The state of the art in the field of specific learning disabilities and its associated fields is such that it is not presently possible to specify exactly all of the components of each specific learning disability. There remain strong opposing professional opinions as to the validity of a specific learning disability. At present, the only generally accepted manifestation of a specific learning disability is that there is a major discrepancy between expected achievement and ability which is not the result of other known and generally accepted handicapping conditions or circumstances.

2. There exists no hard research data collected on a large enough sample in order to state, with certainty, which are the common characteristics of all learning disabled children.

3. There are several theories as to what causes children to have specific learning disabilities.

4. There appear to be no generally accepted diagnostic instruments presently available which can be singly and appropriately utilized with all children with a specific learning disability.

This analysis of the definitional problem is also supported by the findings of the General Accounting Office (Fogel, R.L., et al, 1976). They reported that studies reviewed showed very wide variations of definitions and, as a result, they developed an operational definition of the ability-achievement discrepancy type.

Ideally, under these circumstances one should collect, from a representative sample, data descriptive of several plausible definitions; apply construct validity analyses to these data, from these results determine the definitions best supported empirically, and then collect a new set of data from a new sample to determine incidence rates and provide remedial recommendation. The time and resources available for this study did not permit such an approach. Therefore, an a priori definition, based upon profile discrepancy in learning and ability, has been used to select a measurement package, and thereby to determine the incidence of learning disability in terms of such a definition. Subsequent analysis of the collected data is expected to provide the opportunity to refine the definition empirically, thus adding to the accumulated knowledge about learning disability, and if resources permit, making it possible to reanalyze the collected protocols for revision of the incidence estimates.

The second constraint was imposed by the impracticality of bringing the students who participated in the study to the testing site on more than one occasion. Individual testing consumes a great deal of professional time, thereby consuming a major portion of the study's resources. It is also common knowledge that a period of two and one-half to three hours is the maximum for a session with a single individual. As a result of this situation, the study was limited to the data available from each student's records and from an assessment session completed in a two-and-one-half- to three-hour period. Within these two constraints

the procedures were designed to produce the maximum of usable information and to review this information in a systematic and replicable manner. The nature of the problem required the exercise of considerable professional judgment, within guidelines which could be applied in a variety of settings.

The details of the procedures are presented in the following pages.

II. THEORETICAL APPROACH TO THE ASSESSMENT OF LEARNING DISABILITIES

The procedure described here represents an approach necessitated by study priorities. An early start for the treatment phase of the project was imperative. The approach of providing a priori definition and decision rules for diagnosis seemed the only way to provide potential treatment of subjects on schedule. The data considered must, of course, be tightly focused on the a priori definition rather than broadly focused in more general domains. Second, because individual diagnostic decisions had to be made as soon as possible following data collection, it was necessary that all data be in an immediately meaningful form. That is to say that we did not have the time to collect and analyze group performance data necessary to providing normative and comparable meaning to the variables assessed. For that reason, to make immediate decisions possible, we were limited to the use of presently normed instruments.

Variable Selection

A major problem in assessing learning disabilities (LD) is that of selecting a reference system or systems of learning-related behavior. Such systems are comprised of specified constructs — traits or types of behavior — and specified interrelationships among them. Two types of reference systems are widely reflected in the LD literature: clinical approaches consisting of numerous, often unspecified patterns of responses; and a symptomatic typological approach consisting of relatively well defined components of LD -- aphasia, dyslexia, hyperkinesis, and so on. Neither approach seemed adequate for this study -- the former being largely idiosyncratic to the particular clinician and unreplicable, and the latter casting a net that is agreed upon as too narrow and somewhat superficial. Most educators would agree that LD consists of more than aphasia, dyslexia, and

hyperkinesis and, furthermore, that it is the shared characteristics of the syndromes rather than their surface symptoms that are of interest to a complete LD conceptualization. Rather than accepting either the weaknesses of the clinical/judgmental or symptomatic/partial definitions, we have chosen variables according to a more basic psychological reference system and have selected measures (and records of scores) which are intended to reflect clinical and symptomatic/typological points of view as well.

Within this approach, we had to settle for an incomplete solution since testing time had to be held to reasonable limits and available information from records was only marginally helpful. However, we proposed the allocation of a small portion of testing time to gathering data on the adequacy of the reference systems and measures employed.

In the following sections the reader will notice "marker variables." These short research instruments were intended as checks on the structure of the diagnostic instruments and were not intended for use in making diagnostic decisions. Correlations between marker variables and diagnostic indices are being studied for convergence and divergence, and factor analyses will be performed to confirm or deny the existence of the intended structure in the diagnostic battery. This use of scarce testing time seemed appropriate due to the admittedly special nature of the LD and JD populations. Though we may be well assured that we understand the meaning of performance levels on digit span tasks in the normal population, for example, we were far less assured regarding their meanings in special groups. The marker tests enhance our ability to provide ecologically valid interpretations of performance and to describe the structure of traits in the study's potentially unique populations.

Instrument Selection

The measures listed in the following chart reflect several approaches to LD. The theoretical work of Thurstone, Guilford, Cattell, and Witkin on the structure of abilities guided our thinking. The factor analytic studies of Witkin and his colleagues (Witkin, Dyk, Faterson, Goodenough, & Karp, 1974) were deemed especially relevant, as they provided common findings with the various theoretical structures, and their use of the WISC provided the desired link with a widely used clinical instrument. In addition, clinical conceptions of LD were reflected in the choice of the Bender-Gestalt and other perceptual measures.

The WISC-R is so widely used and well known that a detailed description is not needed here. In addition to theoretical considerations (reflected in the additional scoring for the Witkin, et al factors), the availability of WISC-R scores allows for direct comparisons between results obtained in this exploratory study and those obtained in many other studies of LD children. Moreover, the significance of discrepancies between any two WISC-R subtest scores or between the Verbal and Performances Scales is well established in the literature, thus falling easily into the pattern or discrepancy model used here.

In addition to the conventional WISC-R scoring procedures, we obtained scores following Witkin, et al (1974) that yielded indices of the following three intellectual factors: Analytic Functioning, Verbal Comprehension, and Attention/ Concentration. Four short instruments were used to mark two of the factor scores. The Analytic Functioning score was marked by results on the Hidden Patterns Test (CF-2) and the Children's Embedded-Figures Test (CEFT).

The Number Comparison Test (P-2) and Identical Pictures Test (P-3) were used to mark the Attention/Concentration factor. Because of the robustness of the Verbal Comprehension factor and because many of the achievement measures are so infused with this factor, no compelling need was seen to provide additional markers of Verbal Comprehension.

The Bender-Gestalt is another widely used, well known instrument that needs no description here. It, too, will allow for comparisons between results obtained in the current study and the LD literature. It should be noted, however, that the Bender is suited more to the perceptual ability development of younger children than those in the current study, and the results may therefore be most useful diagnostically as a "low pass" screening device. The scoring method chosen was that developed by Koppitz (1964) with a threshold of four or more errors. The Thurstone Flags and the Swinton-Wepman Test of Memory for Visual Orientation provided markers of the Perceptual Ability factor as assessed by the Bender-Gestalt Test.

Both the Woodcock Reading Mastery Test and Keymath Diagnostic Arithmetric Test were selected because they provided uniquely precise measurement in a short testing time. Each was developed using the Rasch model, such that there is a common scale on which ability in a specific area and the difficulty of specific items can be represented. As in the WISC-R, all subjects began with very simple items and continued until they responded incorrectly to a given number of consecutive items which have been scaled and ordered according to difficulty.

In view of the established association between hyperkinesis and impulsivity, indicators of impulsivity were assessed through the technique of error-scoring on three instruments: the Coding subtest from the WISC-R; the Number Comparison

Test; and the Identical Pictures Test. This technique produces two scores: the total number correct, and a score representing the difference between the number correct and the number of errors. It is, at best, a soft sign which must be used in conjunction with observed behaviors and other data from the battery or the student's record to serve as construct validation data.

There are, of course, myriad other instruments that were considered and that might have been selected for this study. Those chosen for the assessment battery are integral components of the specific reference system established by Witkin, et al (1974) and are represented in a large portion of the clinical literature on LD.

Diagnostic Decisions and Incidence Estimates

Judgments of LD in the adjudicated JD samples had to made immediately following individual assessment so that LD subjects could be assigned to the experimental conditions of the ACLD remediation effort, which was to start as soon as possible.

It was therefore necessary to consider a priori models that describe LD in terms of levels of performance on the single variables selected, in terms of relative levels -- discrepancies --, or in terms of more complicated patterns involving relative performance or performance levels on numerous selected variables. The selection of such an a priori model was neither easy nor rewarding, for there were myriad proposed candidates ranging from simple ratios or differences between two variables to the detection of complex "within test" patterns of behavior, apparently obscured from those other than the numerous author-clinicians writing about them. The probability of finding a reasonably correct, supportable, or popular a priori definition was either terribly small or incalculable; but we had to provide diagnoses that were made according to specified, replicable decision

rules. Furthermore, these rules had to be sufficiently uncomplicated so that some assurance could be maintained in their proper application under field conditions in the hands of trained, supervised staff. The rules had to accommodate both records and individual assessment data.

We are all aware that the term "LD" is general, ambiguous, variously defined, and rarely expressed in terms of a replicable decision rule. Lerner (1971) records the following variations:

1. Neurological dysfunction or brain impairment.
2. Uneven growth patterns.
3. Difficulty in academic and learning tasks.
4. Discrepancy between achievement and potentiality.
5. Definition by exclusion.

Also, Murray, et al (1976) have provided another definition as follows:

> 1. The diagnosis should be based on evidence which cannot as easily be interpreted as a manifestation of mental retardation, physical handicap, emotional disturbance, or environmental disadvantagement. This does not mean that each individual indicator must be unambiguous, but that the diagnosis should be based on triangulated measures which permit a pattern that is inconsistent with the alternative explanations.
>
> 2. The diagnosis should be accompanied by evidence that a discrepancy exists between achievement and expectation. For example, that a child may be demonstrated to occasionally reverse letters does not constitute a learning disability if the child is reading and writing at the level expected of that age and intelligence.

Finally, the proposed definition of the Bureau of Education for the Handicapped,

$$\text{Severe Discrepancy} = \text{C.A.} \left(\frac{\text{IQ}}{300} + 0.17\right) - 2.5$$

has stimulated intensive discussion (Federal Register, 1976).

While all of these definitions include the idea of a discrepancy, none except the BEH formula can be applied without resorting to rather extensive clinical or psychoeducational judgment. The BEH formula, however, does not take into account possible additional evidence which, considered with more moderate discrepancy than the formula requires, might indeed identify children with serious learning disabilities. For the purposes of this study, we used a definition which includes the principles suggested by Murray, et al and extends them through specification of replicable decision rules. The definition accepts intact the first part of the Murray's two-part definition and specifies the evidence required for possible diagnosis.

In the following sections specific decision rules for evaluating the evidence are described. They were based upon level of performance, in some instances. In record review, for example, JD's with academic achievement commensurate with their ages were removed from further consideration as potentially learning disabled. Also, we recognize certain high ability LD youths might be functioning at reasonable levels of academic achievement. However, specifications for this study did not provide for their inclusion. In those instances where level was of importance, measurement error was specifically taken into account. A second type of decision related to discrepant performance on two or more variables. Discrepancy-based decisions predominated in the treatment of assessment data and took into account both the statistical and the substantive significance (educational, psychological) of the magnitudes of differences. The statistical treatment was imperfect, however, for the intercorrelations among measures used were unknown. Finally, data from perceptual tasks were used in a confirmatory sense because adequate normative information on the measures selected was not available; and we were not able, therefore, to transform scores to a scale comparable to other measures.

Specifically, the Bender-Gestalt was used in conjunction with ability and achievement levels or discrepancies to affirm or deny perceptual involvement in apparent difficulties.

Discussion and specifications in the following "Process" sections provide decision rules in precise detail. A preview here might point out that a delinquent who was judged LD was not mentally retarded, nor severely emotionally disturbed, nor achieving at an age- and experience-appropriate level. His performance on assessment measures showed discrepancies at least as large as the difference between the means of age groups two years apart. Observation during the assessment provided some assurance regarding the youth's vision, hearing, and freedom from obvious physical handicap, as well as some data on potential behavioral correlates of LD and test-taking behavior which assisted in test interpretation. This seemed to us to be a reasonably large information yield, given our limits of testing time.

III: PROCESS

Records Review

After sample selection, sets of actual records were used for training the records reviewers. The review process involved, first, the recording of relevant data on subsequently described forms and then followed a systematic decision process, assigning each record to an appropriate category. The review resulted in three categories of decision: those youths who were screened from further consideration because of mental retardation or primary emotional disturbance; those who were interviewed; and those who were diagnostically assessed. The branching of the review process at various decision points is illustrated by Figure I. Decision points are numbered on this figure.* As mentioned previously, two prior decisions applied to the public school sample and one to the adjudicated delinquent sample. The decision common to both samples was age appropriateness. The records of youths born after March 1, 1965 and on or prior to March 1, 1960 were eliminated from further review.** For the public school sample, youths with records of adjudicated delinquency were cross checked against court records and, if appropriate, included in the delinquent sample. Also, designation of visual, hearing or speech problems in the records were noted on the basic data form for possible consideration at a later date. In some cases records were inadequate for the record reviewer decisions. These were categorized for diagnostic assessment (DA).

---

*The numbers are not in strict ordered sequence because branching and recycling prevent such numerical representation.

**The study was originally designed for 12 to 15 year olds. The sample was expanded to include 16-year-olds and youths who had just turned 17-years-old in the record review process.

The first decision rule required consideration of a recorded clinical diagnosis of learning disability (LD). If such a judgment was in the record, the youth was assigned to the category for DA. If it was absent, the second decision rule concerned a diagnosis of mental retardation (MR). We selected the cut-off score of 75 on the most recent psychological examination for this determination, with the further provision that if a discrepancy within the psychological test profile of greater than 9 points existed, students with composite scores of 68 or higher were considered further. If the test used in the determination of mental retardation did not provide part scores, the record was examined for achievement profiles which were unusually mixed or other scores such as the Bender-Gestalt, which had been interpreted as suggesting perceptual and/or general learning problems. When these were found, the record was subjected to further review. Otherwise, a judgment of MR was made and the record was categorized as screened out of the study.

The third decision rule applied to emotional disturbance (ED). If there was no record of ED, the record received further screening. If an ED diagnosis was present, it was evaluated for severity. There is considerable uncertainty about the primacy of emotional disturbance in learning disabled students. This study, therefore, included students so diagnosed unless there was evidence to suggest the impossibility of securing valid information from the student because of profound disturbance. For both JD and non-JD students, if the ED diagnostic report contained in the record characterized the student as withdrawn, unable to relate to the examiner, generally untestable or presenting unusually bizarre or severe symptoms, including a record of hospitalization for these symptoms, the student was categorized as screened out of the study. If the behaviors, on the other hand, were more passive, such as shoplifting, taking small amounts of money

from parents, lying or truancy, the record was reviewed further. The diagnostic supervisor made a clinical judgment on each case which might be ruled out because of ED.

For those records subjected to further review, the fourth decision rule considered achievement scores. Where these were available, the levels of achievement scores were considered. Records of achievement scores which were at or above a point two standard errors of measurement below the mid-point score for children of similar school experience were placed in the interview category. All others were reviewed further.

Decision rule five concerned profile discrepancies, both between ability and achievement scores and within achievement score profiles. While it would have been desirable to determine the significance of a profile difference through the use of a standard error of a difference formula, the variety of tests occurring in the records and the lack of certain necessary statistics for these tests rendered such formulae impractical. In particular, the intercorrelations between all possible pairs of ability and achievements tests were unavailable. The possible standard error bands therefore had to rely on formulae which did not take into account intertest correlation, but required only the variances and reliabilities. The formula used is available in many standard texts. Its application required conversion to a common scale, under the assumption of a normal distribution.

The particular version was adapted from Stanley's (1971) formula 21. The signficance of differences was established by this method, but the method did not determine the decision. Rather, the width of the band was set by what we call "diagnostic significance." In the case of the ability/achievement discrepancy, this value was determined by a score spread which represented test performance equivalent to two years' difference. Two years was selected because the perfor-

mance of urban populations, heavily represented among adjudicated delinquents, tends to be, in general, about one year below the norm on the average. This difference can be reasonably attributed to environmental factors, and consequently must be allowed for in reviewing discrepancies. When a significance test was applied to these differences, using equivalent scores and reliabilities from, for example, the Stanford Achievement Test-Advanced Battery, we found that the differences were reasonably unlikely to be artifacts of the testing.

To repeat, decision rule five required that if the achievement score differed by the equivalent of two years from the ability score on one or more tests, the record was placed in the category for diagnostic assessment. A mixed achievement profile with two-year discrepancies also resulted in classification for diagnostic assessment. If, on the other hand, no such difference was discovered, the record was reviewed for evidence of hyperactivity, unusually illegible handwriting, perceptual or motor performance tests which indicated malfunction and, if grades were available, for erratic grade profiles (decision rule 6). Also significant for this rule were abrupt changes in profile character over time. That is, achievement might have been reasonably consistent through the early grades, but become erratic in the later grades. If any of these were present, the student's record was placed in the category for diagnostic assessment. Otherwise the record was placed in the category for interviewing.

For those records which showed an achievement score profile but no ability score, decision rule four (level of achievement) was applied. Records falling within the achievement limits were considered for grade profile if grades were available (decision rule 7). A profile which was above average was categorized for interview. If the profile was average or below, the grade profile character was considered (decision rule 8). A mixed profile was defined as one which

contained grades in reading, language, math, and science which, for two separate years, showed deviations of at least two letter grade intervals. For example, a letter grade profile CCBD was considered mixed, whereas CDCC was not. Other marking systems were converted to a five-interval scale, with the failing point being considered the top of the lowest interval, and used in an analogous manner. If the institution assigning the grades reported its own interval scale in a usable form, these intervals were applied. When a mixed profile was present for two years, the student was assigned to the category for diagnostic assessment. For those records which did not show a mixed profile, decision rule six was applied. In records for which grades were not available, rule six was applied in terms of the indicators other than grades.

In those cases where ability scores but not achievement scores were available, the availability of grades was considered. If they were available, the categorizing decision was made in terms of decision rules seven, eight, and six, as applicable in the rule-determined sequence. Where grades were not available, decision rule six was used to assign the category. The same decision process applied when neither achievement nor ability scores were available. It should be noted that in no instance was a student screened out of the diagnostic assessment category on the basis of a group ability measure alone. It should also be noted that decision rule six was applied as the final test when any other data suggesting learning disability were lacking. Whenever insufficient data were present to make a decision on these rules, the student was diagnostically assessed.

The entire records review process was under the direction of the diagnostic supervisor, a certified diagnostician. This person paid particular

attention to the decisions about emotional disturbance, personally reviewing clinically those cases where the basis of the recorded diagnosis was not obvious. The interview process is described next.

Interviews

Interviews were conducted following the records review process for all students except those who were not previously screened out because of evidence of mental retardation or emotional disturbance. Students were interviewed whose records indicated (1) high achievement scores, (2) consistent high grades, or (3) non-discrepant academic profiles with no other LD indicators. Students who were diagnostically assessed were interviewed immediately following testing. Interviews were recorded on special forms incorporating content provided by Creighton Institute. The general purpose of the interview was to examine whether students with learning problems get into more trouble than students without learning problems, as well as to collect pertinent background information about the participants.

The first phase of the interviewing process involved training the interviewers. The preliminary part of this training, which was also prerequisite to ensuring successful interview experiences, involved learning how to establish rapport with the students. The diagnostic supervisor trained interviewers in techniques of establishing rapport, which included such topics as (1) general characteristics of the adolescent (social, emotional, physical), (2) potential psychological problems inherent in the adolescent/adult interaction, and (3) desirable interviewer characteristics (e.g., voice intonation, language usage level, eye contact, authoritative attitude level). Part of the interviewers' training involved participation in hypothetical interview situations from the

position of either the interviewer or the student, and responding verbally about how one would deal with a stated problem which might be encountered during an interview. In particular, this training stressed the importance of ensuring the student that confidentiality would be protected.

Each interviewer spent approximately 20 minutes with each student in the actual interview process. This 20-minute block of time for interviewing was divided as follows using interview content provided by Creighton Institute:

	Time (minutes)
Rapport building	5
Other designated interview schedule components	15
	20

At the completion of the interview, the interviewer reinforced the positive intent of the situation by thanking the student for his assistance, attention, and cooperation. Completed interview forms were collected by the diagnostic supervisor, and the data were later forwarded with the diagnostic data to Creighton Institute.

Diagnostic Assessment and Decision Process

The role of the diagnostic assessor, although seemingly limited because of its very specific performance guidelines, was vital to the collection, recording and transmission of accurate data. Performing this role successfully subsequently required such things as:

1. Working familiarity with each test in the battery.
2. Adherence to standardization requirements, with sensitivity to the population being tested (e.g., pronunciation differences).
3. Accurate timing, where required.
4. Consistency in verbal feedback to student responses.

5. Maintaining eye contact with the student.

6. Preparing test materials prior to meeting the student.

Specific instructions included the following admonitions:

> Remember that adolescents in general, and particularly institutionalized juvenile delinquents and learning disabled youth, may tend to give up easily and thus may need continuous encouragement. You should also be aware that certain questions on the WISC-R may be threatening in content or presentation form to adjudicated delinquents. Specifically, we suggest asking "what do people say a 'thief' is?" in the Vocabulary subtest, and in the Comprehension subtest, we also suggest introducing questions 4 and 9 with "what are some reasons that people say ...". It would be well to remember, too, to keep response sheets and manuals with visible answers from the open view of the student during testing. Other than for these minor modifications, the test administration and scoring directions should be followed verbatim.
>
> Also included in your role as a diagnostic assessor is the administration of a short interview form. This instrument, the content of which has been designated by the Creighton Institute, has been modified in an attempt to reflect contemporary wording, values, etc., but cannot be changed in content. Be sure to fill in the student's Identification Code number at the top of the form before beginning. Responses will be marked directly on the form by filling in the blank, circling the number of correct responses, or placing a check or X under or beside the response given. The introductory remarks should be read verbatim but can be expanded slightly to facilitate cooperation and trust, but exact item wording should be presented verbatim, if possible.
>
> As with the test instruments, it is to your advantage to be familiar with the items on the interview guide prior to administration so as to be able to maintain a relaxed but sincere atmosphere. You should remember to thank the student for his cooperation, assistance and patience at the completion of the assessment/interview session.

The diagnostic assessment sessions began with introductions and rapport building activities similar to those utilized for the interviews as described in part A of this section. The tests to be administered included the WISC-R, following standard procedures, followed by an achievement test in reading and in math, the Bender-Gestalt, Rosner's Auditory Analysis Test, and six short data collection devices which were selected to provide corroborative and analytical information in support of three major components of the assessment. The interview questionnaire completed the data collection.

The tests were administered in the order presented below to vary the nature of the type of tasks introduced to the student and, hopefully, to increase his willingness to cooperate over the three hour testing and interviewing session. Specifically, the order in which the battery was presented was:

1. WISC-R
2. Bender Visual Motor Gestalt Test
3. Woodcock Reading Mastery Tet
4. Rosner's Auditory Analysis
5. Hidden Figures Test
6. Key May Diagnostic Arithmetic Test
7. Children's Embedded Figures Test (Part 2)
8. Number Comparison Test
9. Hidden Patterns
10. Swinton-Wepman Visual Orientation Test
11. Thurstone Flags

After the session, the student then returned to his school or home.

When the student had left, the diagnostic assessor entered his or her observations of behavioral indicators on the records form. The behaviors of interest are shown in Figure II. They included four observations specific to the WISC-R. The diagnostic assessor then scored each of the test answer sheets and entered the results on the basic data form.

Finally, using the decision rules discussed in Section II, the diagnostic assessor recorded a coded judgment of LD/not LD on the Basic Data Form. These decision rules required the conversion of the scores of each test to a common scale. Under the assumption of a normal distribution, and given knowledge of the mean and standard deviation of each test, a direct conversion to a standard scale was made. In the interest of convenience, the "T" scale, with a mean of 50 and a standard deviation of 10, was selected. Conversion tables were provided for direct look up and entry on a profile form.

The first consideration in the LD/not LD judgment was a review of the profiles for discrepancies at least equivalent to the difference in group means

of groups two years apart. These differences may have occurred within the score patterns of the WISC-R, including the Witkin factor scores, between the WISC-R scores and any or all of the achievement scores, or between the achievement scores. When discrepancies of the indicated magnitude existed among all three sources of score data, the decision was clearly LD. When two sources showed discrepancies, the same decision applied if there was any supporting evidence from the Bender or the observations. When only one pair of scores showed discrepancy, supportive evidence from the Bender and two or more pronounced characteristics from the behavioral observations were required for LD diagnosis. Cases which showed no significant discrepancies, as defined, but demonstrated two years below level achievement and included observations of difficulty in following oral directions, motor difficulty, paper rotation, productive language problems, distractability, and at least one of the WISC observations were also judged LD if the full scale WISC-R score was at least 75. Cases which did not meet any of these criteria were judged non-LD. In order to maintain the greatest degree of independence in subsequent judgments of the protocol, each diagnostic assessor was provided with a sealed envelope containing a positive and negative LD code number for entry on the form. During training careful emphasis was placed upon the importance of attaining independent judgments. The code assignments were retained in Princeton, and a periodic check of their consistent application was conducted by persons not otherwise involved in the diagnostic assessment.

After the Basic Data Form was completed, it was given to the Diagnostic Supervisor for secure storage until the first data review was conducted. This review occurred in general on every fifth day of assessment. Each diagnostic assessor was given a set of data forms to review for students assessed by others. The decision rules were applied and a judgment of LD/not LD was made by the reviewing assessor.

It should be noted at this point that in some cases the decision rules did not provide an unambiguous conclusion. In these cases the diagnostic supervisors were instructed to resolve the situation with a clinical judgment based on all the relevant data which were recorded on the form.

When this final judgment was recorded, the basic data form was sent to ETS Princeton for transcribing onto tape and for random selection into the remediation and comparison groups of those adjudicated youth who were judged LD. The records of the remediation candidates were then reviewed by the remediation diagnosticians, and copies were transmitted to the ACLD remediation teams, accompanied by the remediation recommendations.

Quality control checks were spaced such that each diagnostic assessor received a check on the average of every eighth day. The checks were conducted by the diagnostic supervisor, who spent the entire assessment session with the diagnostic assessor. To reduce the impact of two adults working with one student, the diagnostic supervisor greeted the student, escorted him to the assessment room, and introduced the diagnostic assessor. The supervisor then told the student that both adults would work with him during the session, and occasionally contributed supportive comments. Sufficient time was provided in the schedule for occasional group discussion of the assessment process.

IV. ADMINISTRATIVE ARRANGEMENTS

Major Tasks and Staffing

Following sample selection, the implementation of this study required completion of several main tasks. They were: review of school and court records of students in the respective samples for evidence of possible LD; interviews to collect background and experience data; diagnostic assessments to provide the information to confirm or disconfirm the possible LD hypotheses indicated by the records review; review and decision based on these data; analysis of records and assessment data to provide remediation suggestions to the ACLD remediation team; and assessment of the remediation and comparison samples for post-treatment data. These tasks were accomplished by teams of professionals under the direction of diagnostic supervisors, as will be the posttesting after the completion of the remediation program.

The diagnostic supervisors at all three sites were practicing certified psychologists. The supervisor of the Baltimore site also served the function of Associate Project Director. In addition to the required credentials and active engagement criteria, the supervisors also met the qualification of being experienced in working with adjudicated delinquents.

The diagnosticians had equivalent qualifications to those of the supervisors. In addition, experience in educating LD children was expected of this group. The primary role of the diagnosticians was to review the test protocols and the BDF's for purposes of providing recommendations to the ACLD remediation team. The reviewers, interviewers and diagnostic assessors were people with degrees in psychology, education or similar fields and experience in working with educational records and with adolescents. Figure II describes the qualifications and responsibilities of every member of the staff.

Throughout most of the pre-test phase, we found that a team of five record reviewers and eight diagnostic assessors was kept busy at all times. In most cases the record reviewers also served as interviewers although an additional one or two staff members served in the interview function.

Training

Training sessions were most significant as procedures for assuring uniform data collection. The records reviewers/interviewers and diagnostic assessors were trained on-site by the site diagnostic supervisors and by ETS staff.

Training began with a thorough review of materials and procedures. The records reviewers/interviewers were trained first, since their work began almost immediately. The record reviewers administered the interview form to one another and performed dual record reviews as part of this procedure. Following this, the diagnostic assessors (DA) were trained and a program of trial assessments among the DAs was conducted. Each DA administered and was tested with the diagnostic assessment package. Finally, a period of testing by a pair of diagnostic assessors with selected cases was conducted at the outset of the diagnostic period. (This served as an initial quality control check; quality control is further discussed in the next section.) A final crucial element in the training, interspersed throughout its course, involved discussing suggestions, anecdotes, and procedures which were useful in working with adolescents. These were provided by the experience of the diagnostic supervisors and our consultants.

Quality Control

Any data collection which is as extensive and complex as that designed for this project requires careful attention to assure the quality of the data. Several procedures were designed to accomplish this purpose. First, 5 to 6%

of all diagnoses (spaced across the entire period of assessment) were quality controlled by the diagnostic supervisor; this was accomplished by the actual presence of the supervisor during the required number of diagnostic assessment.

Second, a one-half-hour protocol check of each diagnostic assessment was made in a cooperative effort between the diagnostic supervisor and the diagnostic assessors. (The assessors did not check cases which they themselves originally assessed.) This protocol check consisted of a review of the entire data form, including school or court data transcribed thereon, assessment data, the application of the decision rules, the assessor's comments and the recorded observations made by the diagnostic assessor who collected the data. This review lead to an independent judgment of LD or not LD by the reviewer. The first review was "blind" because the judgment of the data collector was entered in code. The diagnostic supervisors reviewed the two DA decisions to identify and resolve any difference in judgments.

Third, a second protocol check of the LD/JD diagnosed cases at each site was made by the diagnosticians. This check was part of the review process which produced remediation recommendations.

Finally, periodic site visits by a two- or three-person monitoring team (members of the advisory committee) were conducted during the diagnostic testing period for purposes of reducing variation in diagnostic points of view. The monitoring team observed actual assessments taking place at each site.

The team members were J. Richard Harsh, a clinical psychologist who has both worked in California's Juvenile Halls and has conducted extensive research with another team member, Professor Nadine Lambert, of UCLA Berkeley.

There were also cross checks of samples of deidentified protocols across sites, followed by an analysis of these cases to achieve greater commonality and to check for diagnostic drift. This entire process can be classified as an

additional quality control check. It had been suggested that external evaluators conduct this final review. However, its prime purpose was to facilitate cross-site uniformity by providing common sets of data for discussion. The direct resolution of differences of opinion among the diagnostic supervisors was a part of this process.

Forms

1. Basic Data Form.

All data gathered throughout the record screening and diagnostic assessment processes were recorded on an individual form, which, with the attendant answer sheet(s), provided a complete diagnostic package. (A separate form was used by the interviewer. It was later merged with the records data.)

Part I of the form was used to enter information found in the subject's school and court records. This information provided the tool for decision making in response to screening criteria as they were outlined in the decision tree (see III. Process — A. Records Review). These criteria included prior diagnoses of physical handicap, emotional disturbance, mental retardation, or LD, as well as ability and achievement data, recorded grades, profile character, and comments regarding LD symptoms. Also included in Part I were spaces for recording speech, vision, and hearing problems

Part II was used to record the student's performance on the testing instruments. The test instruments were hand-scored by the DA to produce this record. The DA entered a coded judgment of his/her opinion as to whether the boy had a learning disability.

Part III of the form was used by the diagnostic assessor for written observations of each student's behavior throughout the assessment. Provisions were made for comments by the diagnostic assessor or the supervisor.

Part IV of the document contained the overall judgment of a second diagnostic assessor and/or the diagnostic supervisor. This summary was based upon a half hour protocol check and resulted in either a "non-LD" or "LD" decision. Also included in Part IV was an indication of current detention status, i.e., none, probation, parole, or institutionalized.

2. Profile and Summary Sheets

a. To facilitate comparison of information from several sources, significant data were entered on a profile sheet. The major elements in this profile were the five scores from the WISC-R (Verbal, Performance, Analytic Functioning, Verbal Comprehension and Attention/Concentration) and the achievement scores in reading and mathematics.

b. A summary sheet was used by the diagnostician for all remediation cases who were assessed LD. (Note the exclusion of non-JDs in this final diagnostic process, since none of these cases were included in the remediation sample.) This summary sheet provided handwritten data recorded during an additional protocol check (one hour) made for each youth who was randomly selected for the remediation sample. Its major purpose was to provide recommendations for remediation for the student, and it was included in his diagnostic package as part of the Basic Data Form.

V. APPLICATION

The procedures outlined above were applied at three sites; Baltimore, Indianapolis and Phoenix. The public schools, juvenile centers and juvenile institutions cooperated in each case. At the date of this writing the study is still underway.

Some interesting miscellaneous sidelights might be useful to those who anticipate similar studies:

-- An experienced records reviewer can average about eight records per day. Records are stored in a variety of locations, and are not necessarily intact. Where they are well organized, as many as twelve records per day can be reviewed using these procedures. Review includes recording pertinent data on the Basic Data Form and making the decision to asses or not to assess.

-- The schedules of many schools are such that the completion of two diagnostic assessments per day is frequently impossible. An average of one and one half per day is a reasonable expectation.

-- The only effective way to arrange assessments and interviews with parolees and probationers is through direct personal contact. Ample personnel time for such an approach should be provided.

-- There are remarkably few outright refusals to participate on the part of the students once arrangements are made. Genuineness, warmth and patience on the part of the diagnostic assessors and interviewers probably contribute most to the successful implementation of the procedures.

The collection of these kinds of data in a field operation is a highly person-intensive activity, and will inevitably be a costly process. It is, however, the only available method which is applicable for research on the kinds of questions addressed by this study.

Figure I

Selected Instruments

Intellectual Functioning	Items	Time
WISC-R (Conventional scoring)		
Verbal Scale		
Information	30	
Similarities	17	
Arithmetic	18	
Vocabulary	32	
Comprehension	17	
Digit Span	14	
Verbal IQ		
Performance Scale		
Picture Completion	26	
Picture Arrangement	12	
Block Design	11	
Object Assembly	4	
Coding	45	
Performance IQ		
Full Scale IQ		45-60′ est
WISC-R (Additional Scoring)		
Analytic Functioning		
Block Design	11	
Picture Completion	26	
Object Assembly	4	
Verbal Comprehension		
Vocabulary	32	
Information	30	
Comprehension	17	
Similarities	17	
Attention/Concentration		
Digit Span	14	
Arithmetic	18	
Coding	45	
Factor Markers*		
Hidden Patterns CF-2	400	3′
Children's Embedded Figures Test (CEFT)	14	10′
Number Comparison P-2	96	3′
Identical Pictures P-3	96	3′
Thurstone Flags	21	5′
Test of Memory for Visual Orientation	20	5′

	Items	Time
Perceptual Ability		
Bender-Gestalt	9	15′
Rosner's Auditory Analysis Test	20	10′
Achievement		
Reading		
Woodcock Reading Mastery		
Letter Identification	45	
Word Identification	150	
Word Attack	50	
Word Comprehension	70	
Passage Comprehension	85	
Total Reading		20-30′ est
Arithmetic		
Key Math Diagnostic Arithmetic		
Content:		
Numeration	24	
Fractions	11	
Geometry & Symbols	20	
Operations:		
Additon	15	
Subtraction	14	
Multiplication	11	
Division	10	
Mental Computation	10	
Numerical Reasoning	12	
Applications:		
Word Problems	14	
Missing Elements	7	
Money	15	
Measurement	27	
Time	19	
		30′ est
Additional Observations		
Vision		
Hearing		
Impulsivity		

Total estimated time for entire battery: 141 - 166 minutes.

---

*The standard directions for the marker tests have been modified as necessary.

Figure II

Diagnostic Staff

Title[a]	Responsibilities	Qualifications
Diagnostic Supervisor[b]	Supervises, selects, trains site team Performs quality control Responsible for maintaining security of data at site Contact person for schools and institutions	Practicing certified psychologist Experience in working with adjudicated delinquents Experience in teaching or training of testers
Records Reviewer	Reviews school and court records Enters data in forms Applies decisions rules	Degree in psychology, education, or similar fields and experience in working with adolescents and with educational records
Interviewer	Administers the questionnaire	
Diagnostic Assessor	Administers the assessment package Provides reviews of basic data Administers the posttest	
Diagnostician	Reviews protocols for remedial recommendations to the ACLD	Practicing certified psychologist Experience in working with adjudicated delinquents and with learning disabled children

[a]For study purposes.

[b]It is understood that the Diagnostic Supervisor at the Baltimore site was also the Associate Project Director.

Records Review

Figure III

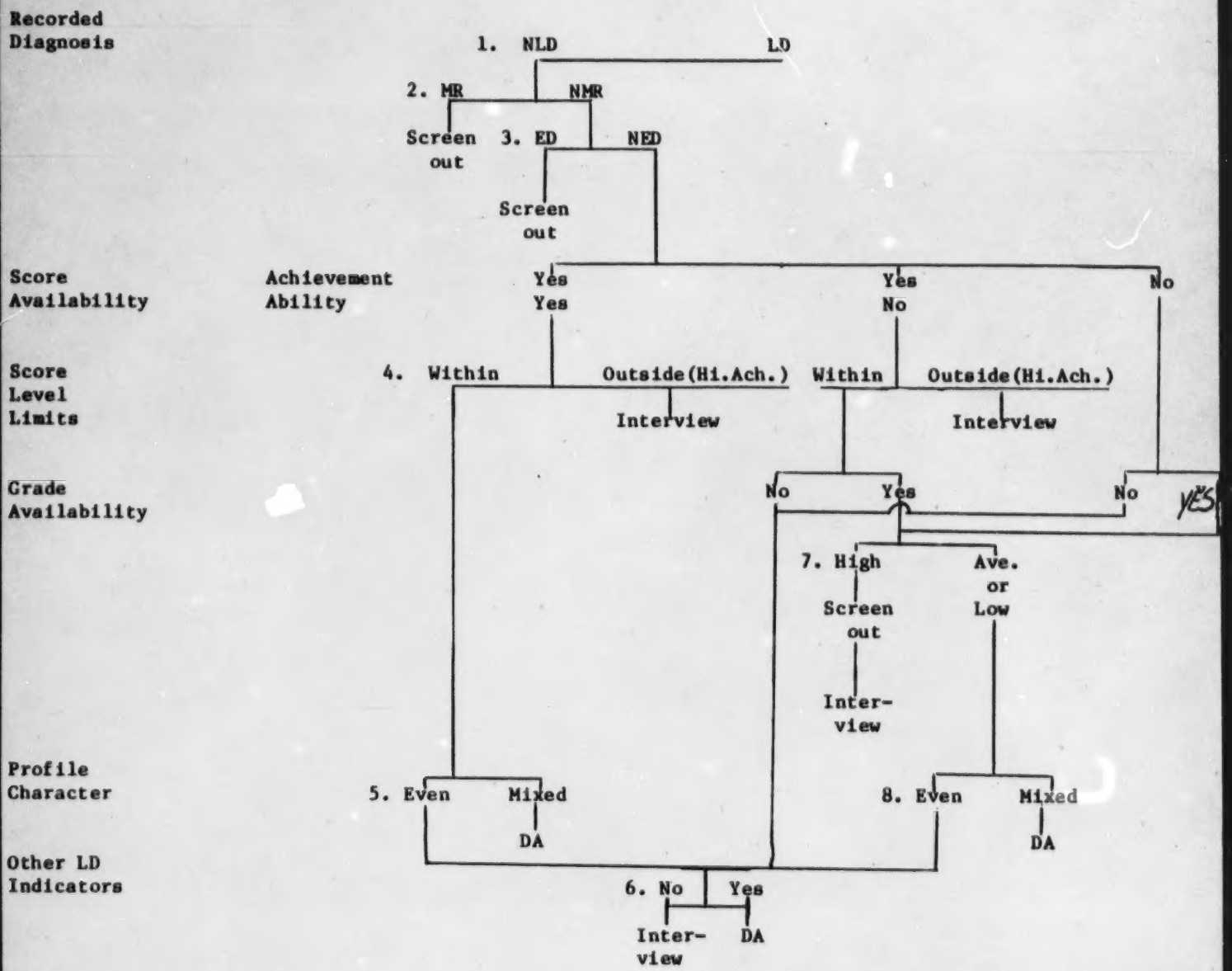

Figure IV

	Pronounced Characteristic	Occasionally Observed Characteristic	Not Observed
Difficulty following oral directions	___	___	___
Low Frustration, e.g., early onset of fidgeting, inattentiveness	___	___	___
Guarded response style (may be withdrawal, hostile response, evasive response)	___	___	___
Repeated verbalization of inability to learn	___	___	___
Gross motor difficulty, e.g., unusual awkwardness	___	___	___
Fine motor difficulty, e.g., difficulty with handling pencil or similar tasks	___	___	___
Manifestation of vision problems, e.g.,squinting, holding books very close, rotation of paper	___	___	___
Manifestation of hearing problems, e.g., favoring one ear, focusing on speakers lips	___	___	___
Continuous rocking, tapping, drumming	___	___	___
Difficulty in Oral Expression. e.g., disjunctive sentences,inconsistent grammatical errors, long latency for common words	___	___	___
Distractability	___	___	___
Additional Comments			
WISC Observations			
Block designs - perseverating in patterns	___	___	___
Inability to perform on sequence test	___	___	___
Inability to complete any math problem	___	___	___
Inability to complete puzzles	___	___	___

References

Federal Register, DHEW, Office of Education, Part III, Education of Handicapped Children - Assistance to States, Proposed Rulemaking, November 29, 1976.

Fogel, R.L., Hunt, J., Baver, R. and Brockway, M. GAO Presentation before the International Conference of the Association for Children with Learning Disabilities on Review of the Relationship of Learning Problems to Juvenile Delinquency. Seattle, Washington, March 1976.

Koppitz, E. The Bender-Gestalt Test for Young Children. New York, 1964.

Murray, C. A. , et al. The link between learning disabilities and juvenile delinquency - current theory and knowledge. American Institutes for Research (prepared for NIJJDP, LEAA). April, 1976.

Stanley, J.C. Reliability. Educational Measurement, Robert Thorndike, (Ed.), American Council on Education, Washington, D.C., 1971.

Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R. and Karp, S.A. Psychological differentiations. New York: John Wiley & Sons, 1974.